Analysis of Multiscale Products for Step Detection and Estimation

Brian M. Sadler, *Member, IEEE*, and
Ananthram Swami, *Senior Member, IEEE*

***Abstract*— We analyze discrete wavelet transform (DWT) multiscale products for detection and estimation of steps. Here the DWT is an over complete approximation to smoothed gradient estimation, with smoothing varied over dyadic scale, as developed by Mallat and Zhong. The multiscale product approach was first proposed by Rosenfeld for edge detection. We develop statistics of the multiscale products, and characterize the resulting non-Gaussian heavy-tailed densities. The results may be applied to edge detection with a false-alarm constraint. The response to impulses, steps, and pulses is also characterized. To facilitate the analysis, we employ a new general closed-form expression for the Cramer–Rao bound (CRB) for discrete-time step-change location estimation. The CRB can incorporate any underlying continuous and differentiable edge model, including an arbitrary number of steps. The CRB analysis also includes sampling phase offset effects and is valid in both additive correlated Gaussian and independent and identically distributed (i.i.d.) non-Gaussian noise. We consider location estimation using multiscale products, and compare results to the appropriate CRB.**

***Index Terms*—Change detection, Cramer–Rao bound, multiscale analysis, step localization, wavelets.**

I. Introduction

Step or edge detection and estimation is a fundamental problem in many areas of signal and image processing that involves basic statistical tradeoffs. A classical linear low-complexity approach to this problem is based on smoothed gradient estimation, also called filtered derivative methods. This approach has been heavily exploited in image processing and elsewhere, e.g., see [10, Ch. 9]. The filter response tends to be localized, so the approach is useful in cases such as images that may be characterized by relatively high signal-to-noise ratio (SNR) and multiple closely spaced change points. Localized response also brings some tolerance to fluctuations in the SNR. Attempting to achieve simultaneous detection and estimation results in a tradeoff between the level of smoothing and the variance of the estimated step location, and this tradeoff is sensitive to the edge shape and SNR. A filtered derivative method that has received a lot of attention is the derivative of Gaussian (dG), which estimates the gradient after smoothing with a Gaussian function. The dG approach can be derived under criteria of detection and localization; see Canny [5], Tagare and deFigueiredo [32], and Koplowitz and Greco [12]. Because of the tradeoffs associated with choice of smoothing scale, it is of interest to study gradient estimation techniques that exploit multiple levels of smoothing (multiscale) simultaneously. Many alternative methods are based on testing for a change in distribution, e.g., see Basseville and Nikiforov [2].

It is well known that wavelets may be used to characterize the local regularity of signals, and may be applied for singularity detection and characterization, e.g., see Mallat and Hwang [17]. In this regard Mallat and Zhong (MZ) [18] developed a discrete wavelet transform based on smoothed gradient estimation, where the level of smoothing varies with (dyadic) scale. This implementation is based on splines, and closely approximates the dG estimator; hereafter we refer to the implementation in [18] as the MZ discrete wavelet transform (MZ-DWT) algorithm. This approach is nonorthogonal so as to preserve regularity information at each point in time for each scale. This has been applied to filtering (denoising) [17], [33], compression [18], and ECG characterization [14].

In the present work we analyze a multiscale analysis technique based on forming multiscale point-wise products of smoothed gradient estimators. This idea was put forward by Rosenfeld and co-workers before the advent of wavelets [23]–[25]. This approach is intended to enhance multiscale peaks due to edges, while suppressing noise, by exploiting the multiscale correlation due to the presence of the desired signal in a direct (albeit nonlinear) way. It is interesting to note that Rosenfeld's original work utilized dyadic scales, both for ease of implementation and because this seemed to work as well as any other combination of scales. The multiscale product idea was recently applied for filtering of magnetic resonance images by Xu *et al*. [33], and for robust shock wave detection in acoustics [26]. An alternative multiscale detection approach is proposed by Denjean and Castanie [7], employing a multivariate test over scales of a continuous-wavelet transform. We consider the step-change problem in additive and multiplicative noise in [31] and [30], including Cramer–Rao bounds (CRB) for location estimation, as well as detection and location estimation via maximum-likelihood, wavelet, and Neyman-Pearson approaches.

We characterize the multiscale product approach statistically and evaluate performance for detection and estimation of edges. We show that multiscale products generally reduce any correlation in the input noise, and that they have heavy tailed distributions. We also characterize the pulse response in the multiscale-product domain. To facilitate the analysis, we use a general discrete-time CRB formulation for both additive Gaussian and non-Gaussian noise, including a closed-form expression not previously available. With this expression, the CRB may be found for any underlying continuous and differentiable edge model, including an arbitrary number of steps. We also show the effects of sampling phase offset τ, for samples of an underlying continuous model. For a single realization, we show that choice of τ can have a significant impact on the CRB.

II. A Nonorthogonal DWT and Multiscale Products

In this section we provide background and a quick overview of the MZ-DWT algorithm. We then introduce the multiscale product and a motivating example.

A. Multiscale Gradient Estimation

Consider a continuous-time wavelet given by $\phi_s(t) = (1/s)\phi(t/s)$, such that $\phi(t)$ meets appropriate criteria to be a wavelet. The continuous-wavelet transform (CWT) of $x(t) \in L^2(R)$ is $W_s x(t) = x(t) * \phi_s(t)$, where $*$ denotes convolution. In this paper we restrict ourselves to 1-D; 2-D extensions are described by Mallat *et al*. [17], [18]. Consider $\phi(t) = du(t)/dt$, with $u(t)$ acting as a local average or smoothing function. Now,

$$W_s x(t) = x * \left(s \frac{du_s}{dt} \right)(t) = s \frac{d}{dt}(x * u_s)(t) \tag{1}$$

provided that $\phi(t)$ is a valid wavelet with zero first moment, so

Manuscript received February 13, 1998; revised September 10, 1998 and December 1, 1998. Part of the results in this paper were presented at the 1998 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-98).

The authors are with the Army Research Laboratory, AMSRL-IS-TA, Adelphi, MD 20783 USA.

Publisher Item Identifier S 0018-9448(99)01995-1.

TABLE I
THE REGION OF SUPPORT (RoS) OF THE MZ-DWT FIR FILTERS AS FUNCTION OF SCALE

Scale	1	2	3	4	5	6	7	8
RoS	2	6	14	30	62	126	254	510

that differentiation and integration may commute [18]. Thus, for appropriate choice of $u(t)$, $W_s x(t)$ can be interpreted as a derivative of a local average of $x(t)$ where the degree of smoothing depends on scale s. Of particular interest here is the case of $u(t)$ closely approximating a Gaussian function; the result is derivative estimation at various levels of smoothing.

The MZ-DWT is a discrete-time approximation to a Gaussian $u(t)$ using a cubic spline (consequently $\phi(t)$ is a quadratic spline). Restricting to dyadic scales, the DWT of $x(n)$, $1 \le n \le N$, consists of

$$W_{2^j} x(n), \qquad j = 1, 2, \cdots, J-1 \tag{2}$$

where $J = \log_2 N$, plus the remaining coarse scale information denoted by $S_J(n)$. This DWT, consisting of $J \times N$ points, is over complete (nonorthogonal). This contrasts with the (perhaps more commonly encountered) orthogonal wavelet transforms where the number of coefficients decreases with scale. The inverse transform may be readily accomplished; see [17] and [18]. In addition, reconstruction from local maxima via alternating projections may also be attained, leading to edge-detection based filtering [6].

To emphasize the linear filtering aspect of (2) and to simplify notation, let

$$y_j(n) = W_{2^j} x(n) = \sum_k h_j(k) x(n-k) \tag{3}$$

denote the DWT response to $x(n)$ at the jth scale, where $h_j(n)$ is the impulse response (IR) of the jth DWT filter. Hereafter we restrict our attention to Mallat and Zhong's implementation (MZ-DWT), although our results are general for a family of linear derivative estimation filters. Here, each filter is an approximation to dG, with smoothing increasing with scale. Implementation details are available in [18]; see also [28] for Matlab code of the forward and inverse MZ-DWT. To avoid discontinuities at the window edges when computing the MZ-DWT, we use an odd-symmetric extension of the data.

The region of support (RoS) of $h_i(n)$ is $\Sigma_{k=1}^{i} 2^k = 2^{i+1} - 2$. These are tabulated in Table I. Also $h_i(n) = -h_i(-n)$, so that $\Sigma_{l_i} h_i(l_i) = 0$, for all i. Note the relatively long RoS for scale ≥ 4, which impacts the use of these scales due to interaction between neighboring edges.

B. Multiscale Products

Working before the advent of the wavelet framework, Rosenfeld and co-workers suggested forming multiscale point-wise products [23]–[25]. This is intended to enhance multiscale peaks due to edges, while suppressing noise, by exploiting the multiscale correlation due to the presence of the desired signal. The multiscale product of the first K DWT scales is given by

$$p_K(n) = \prod_{j=1}^{K} W_{w_2^j} x(n) = \prod_{j=1}^{K} y_j(n). \tag{4}$$

More generally we include an arbitrary set of scales denoted by

$$p(n) = \prod_{j \in \mathcal{K}} W_{2^j} x(n) = \prod_{j \in \mathcal{K}} y_j(n) \tag{5}$$

where $\mathcal{K}$ is an integer set of cardinality K. As further motivation, the number of local maxima due to noise decreases quickly with scale due to the increased smoothing. Specifically, for $x(n)$ white-Gaussian noise, the average number of local maxima at scale $s = 2^{j+1}$ is half that at scale $s = 2^j$ [17]. Thus, while maxima in $W_{2^j} x(n)$ due to edges in $x(n)$ will tend to propagate across scales, the maxima due to noise will tend not to, so that $p(n)$ will tend to reinforce the signal response and not the noise.

An example of the cross-scale product is given in Fig. 1. The signal is $x(n) = As(n) + v(n)$, with $A = 10$, $v(n)$ unit-variance white Gaussian noise, and $s(n)$ taking on values of 1 in the ranges [51, 150], [201, 250], [301, 310], [361, 365], as well as impulses given by $\delta(n-416)$, $\delta(n-442)$, and $\delta(n-445)$. The SNR is 20 dB, defined as $10 \log_{10}(A^2/\sigma_v^2)$. Here, peaks do not align across arbitrarily high scales because neighboring peaks interfere due to lengthening filter responses. The peaks in $p(n)$ are generally well pronounced, except for those corresponding to the isolated impulses between $n = 400$ and 450 in the original time series, where smoothing leads to weakened response at the higher scales. This suppression of impulses and the effect of pulse width is explored in Section IV. Peaks in Fig. 1(f) are generally positive going because of the even number of products, whereas those in Fig. 1(g) are bipolar and preserve the edge up/down direction information. In the following we develop properties and analyze behavior of $p(n)$.

III. PULSE RESPONSE

Next we consider the response to pulses of varying widths, which is important to understand in multistep change applications. The interaction between edges becomes an issue when they are both contained within a filter region of support. Also, the smoothing at higher scales will smear short pulses and dampen the maxima in $p_K(n)$, as depicted in Fig. 1. The suppression of small features in $p_K(n)$ is a direct result of the smoothing at the higher scales, and may be considered beneficial or harmful depending on one's point of view. These effects can be quantified by examining the noise-free response to a pulse of varying width. The left side panel of Fig. 2 shows normalized values of $p_3(0)$, where the input is a pulse of width n_0, given by $s(n) = u(n-1) - u(n-1-n_0)$, for $n_0 = 1, \cdots, 11$. Here $p_3(0)$ is the peak response to the leading edge of the pulse. We plot $10 \log_{10} p_3(0)$, normalized by the maximum of $p_3(0)$ over the range of n_0. Pulses of width $n_0 \ge 6$ yield essentially the same peak response in $p_3(n)$, equivalent to the ideal unit-step response. The case of $n_0 = 1$ corresponds to $s(n) = \delta(n-1)$, an impulse. Fig. 2 indicates a (deterministic) 15-dB suppression of impulses in $p_3(n)$, relative to the noise-free step response which is achieved for a pulse width of $n_0 \ge 6$. The figure also shows $\approx$3 dB suppression of edge response for a pulse of width two (three samples total).

We repeat the above, now using the product of scales 2, 3, and 4, with results shown in the right side panel of Fig. 2. In this case the additional smoothing and omission of the first scale leads to more suppression of impulses relative to the step-response (about 31 dB). However, the price paid is that a pulsewidth of ten or more samples is now required for the edge response to match that of the ideal step-response, and pulses of width five have $\approx$3 dB suppression in edge response.

The impulse suppression effect is dominated by choice of largest scale. For example, using the product of scales 1, 2, 3, and 4 yields essentially identical results as using scales 2, 3, and 4 (right-hand panel of Fig. 2). Use of scales 1 and 2 yields <5 dB suppression of isolated impulses and no pulse leading edge suppression for pulse widths greater than two. These results indicate that the multiscale product has an inherent ability to suppress isolated impulses and narrow pulses, depending on the amount of smoothing utilized.

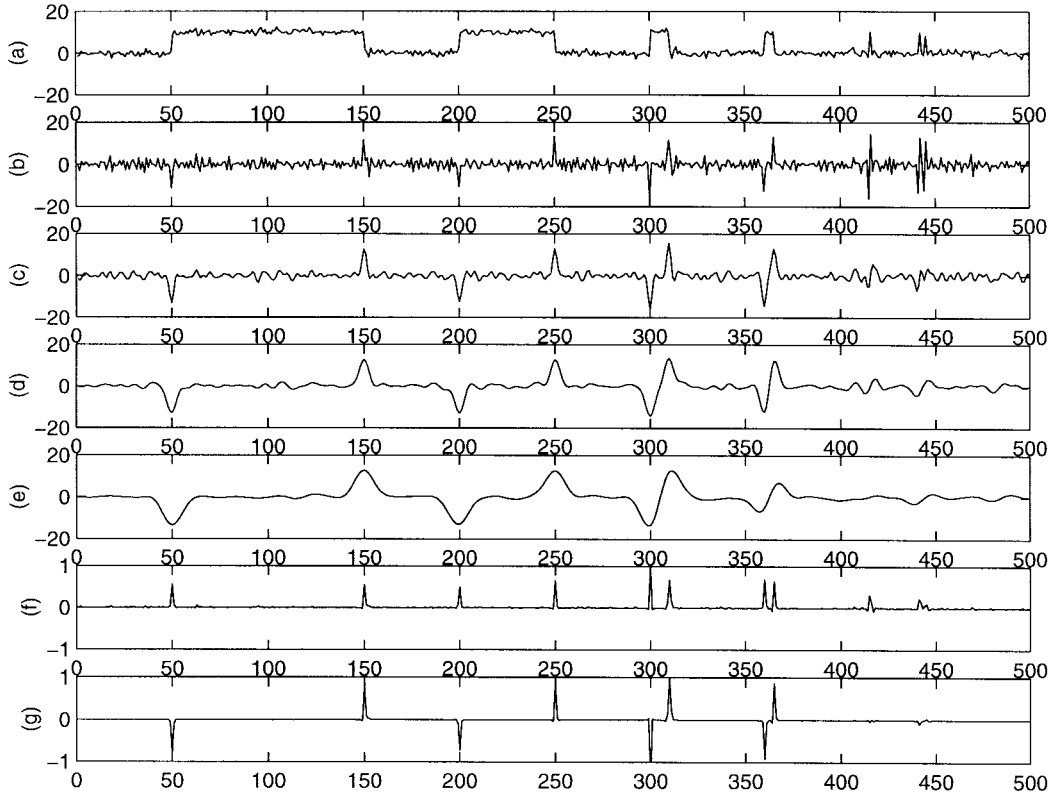

Fig. 1. MZ-DWT example showing (a) time series, (b)–(e) first four scales of the DWT, (f) normalized $p_2(n)$, and (g) normalized $p_3(n)$.

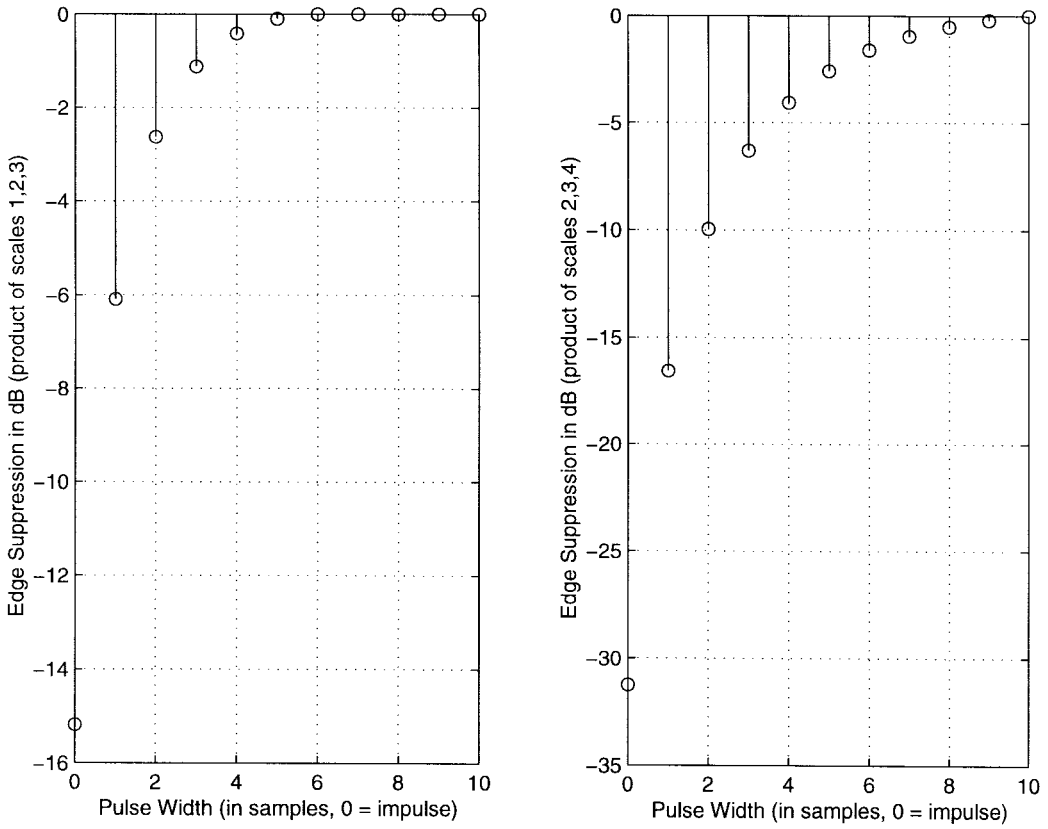

Fig. 2. Pulse leading edge response of the multiscale product, as a function of pulse width. Left panel: product of scales 1, 2, 3. Right panel: product of scales 2, 3, 4.

TABLE II
TOP: CORRELATION COEFFICIENTS ρ_{ij} BETWEEN MZ-DWT OUTPUTS $y_i(n)$ AND $y_j(n)$ (SCALES $s = 2^i$ AND $s = 2^j$), FOR WHITE NOISE INPUT. BOTTOM: MAIN DIAGONAL VALUES OF COVARIANCE MATRIX C, WITH $\sigma_x^2 = 1$

ρ_{ij}						
1.0000	0.5345	0.2097	0.0759	0.0270	0.0096	0.0034
0.5345	1.0000	0.6444	0.2791	0.1074	0.0395	0.0142
0.2097	0.6444	1.0000	0.6787	0.3019	0.1176	0.0435
0.0759	0.2791	0.6787	1.0000	0.6868	0.3075	0.1201
0.0270	0.1074	0.3019	0.6868	1.0000	0.6888	0.3089
0.0096	0.0395	0.1176	0.3075	0.6888	1.0000	0.6893
0.0034	0.0142	0.0435	0.1201	0.3089	0.6893	1.0000

$[C]_{ii}$						
3.5556	1.3951	0.6701	0.3320	0.1673	0.0834	0.0417

TABLE III
ESTIMATED PARAMETERS OF $p_K(n)$ FOR VARYING K, SHOWING HEAVY-TAILED NATURE OF THE PDF'S, AND NORMALIZED SKEWNESS FOR K EVEN

Unit-variance Gaussian input				
K = no. of products	1	2	3	4
mean	0.0000	1.1899	0.0004	0.4948
variance	3.5531	6.3654	10.1739	6.9256
skewness	0.0007	2.4063	0.0143	10.4348
kurtosis	-0.0010	10.1278	52.7261	230.4220
Unit-variance Laplace input				
K = no. of products	1	2	3	4
mean	0.0000	2.3805	0.0029	2.0214
variance	2.6666	5.8274	10.9141	13.3247
skewness	-0.0003	4.0264	-0.1553	18.0487
kurtosis	1.4981	32.0013	171.7648	825.0208

We have also studied these effects using *sums* of MZ-DWT outputs in [30]. Of course, this is equivalent to a single-linear filter that includes the DWT and output summing. In this case, the impulse/edge suppression is dominated by the *lowest* scale terms. We note that impulses are better suppressed by products than sums.

IV. MULTISCALE PRODUCTS ARE HEAVY TAILED

Next we study the univariate pdf's of the multiscale products $p_2(n)$ and $p_3(n)$, for $x(n)$ random Gaussian, and look at detection in the multiscale-product domain. We show that the resulting pdf's are generally non-Gaussian heavy tailed.

A. PDF's of Multiscale Products

Determination of the first-order pdf of $p_K(n)$ is not straightforward, even with $x(n)$ Gaussian, where a closed form is only available for the bivariate case. Let y_1, y_2 be zero-mean jointly Gaussian with covariance matrix

$$C = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{6}$$

where ρ_{12} is the correlation coefficient. The product $z = y_1 y_2$ has pdf [20, Sec. 2.3]

$$f(z) = \frac{1}{\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} e^{\rho_{12}\sigma_1\sigma_2 z} K_0(\sigma_1\sigma_2|z|) \tag{7}$$

which is right-skewed for positive ρ_{12}, and where K_0 is the modified Bessel function of the second kind and order zero.

The correlation coefficients between DWT outputs at scales i and j are readily derived for $x(n)$ zero-mean real-valued white noise with correlation function $E[x(n)x(n+m) = \sigma_x^2\delta(m)]$. We obtain

$$\rho_{ij} = \frac{\sum_k h_i(k)h_j(k)}{\left(\sum_{k_1} h_i^2(k_1)\sum_{k_2} h_j^2(k_2)\right)^{1/2}}. \tag{8}$$

For the MZ-DWT FIR filters ρ_{ij} is shown in the top of Table II for $1 \le i, j \le 7$, and the values on the diagonal of C are shown on the bottom (with $\sigma_x^2 = 1$). Note the decreasing correlation as the scales are separated, with (dyadic) scales a distance 3 apart being essentially uncorrelated. Using Table II we find that for the first two scales $\sigma_1^2 = 3.5556$, $\sigma_2^2 = 1.3951$, and $\rho_{12} = 0.5345$, so that in this case

$$f(z) = \frac{0.5313}{\pi} e^{0.3360z} K_0(0.6286|z|). \tag{9}$$

Fig. 3 depicts pdf's for two cases with $x(n)$ white Gaussian, the theoretical pdf for $K = 2$ from (9), and an unsmoothed histogram estimate for $K = 3$ based on 5×10^6 samples. Normal pdf's with matching variances are shown for reference. Note the skewness in the $K = 2$ case, because $\rho_{12} = 0.5345 > 0$. The positive tail is heavy, showing a strong impulsive nature in $p_2(n)$ on the positive side, with a much lighter tail on the negative side. In the $K = 3$ case the pdf is symmetric with tail behavior eventually heavier than Gaussian.

In Table III we show some estimated statistics of $p_K(n)$ for $1 \le K \le 4$, for independent and identically distributed (i.i.d.) unit-variance Gaussian input and i.i.d. unit-variance Laplace input, including normalized skewness and normalized kurtosis. These estimates were obtained by averaging over 100 runs with 10^5 samples for each run. They demonstrate the heavy-tailed nature of $p_K(n)$, as well as the strong skewness for K even. These may also be evaluated theoretically, although the algebra becomes lengthy as K increases.

An appealing alternative to histogram or kernel-based pdf estimation is to employ an L-term Gaussian mixture model, given by

$$f_m(z) = \sum_{l=1}^{L} \frac{\lambda_l}{(2\pi\sigma_l^2)^{1/2}} \exp\left(-\frac{|z|^2}{2\sigma_l^2}\right) \tag{10}$$

where the model is parameterized by $\Sigma = [\sigma_1^2, \cdots, \sigma_L^2]$ and $\Lambda = [\lambda_1, \cdots, \lambda_{L-1}]$, with $\lambda_L = 1 - \Sigma_{l=1}^{L-1} \lambda_l$. The model encompasses a broad range of symmetric zero-mean pdf's, e.g., see Liporace [15]. Approximate maximum-likelihood (ML) estimates of the parameters are readily obtained using the iterative expectation-maximization (EM) algorithm, e.g., see McLachlan and Krishnan [19, Sec. 2.7]. As is well known, each iteration of the EM algorithm achieves nondecreasing likelihood, although local convergence may be the result, depending on initialization. However, for heavy-tailed symmetric pdfs [all mixture terms zero mean as in (10)] very good results can be obtained for small sample sizes with minimal dependence on initialization. The Gaussian mixture is readily applied to detection and estimation problems in heavy-tailed noise, e.g., see [3], and [13].

As an example, we found $\hat{f}_m(z)$ for $p_3(n)$, obtained from 1000 samples and $L = 10$ in (10). The EM algorithm was iterated 100 times; this was more than adequate in our tests. For initialization we used

$$\begin{aligned} &\lambda_1 = 0.98, \quad \lambda_2 = 0.005 \\ &\lambda_i = 0.0025, \qquad \text{for } 3 \le i \le 6 \\ &\lambda_i = 0.00125, \qquad \text{for } 7 \le i \le 10 \\ &\textstyle\sum = \hat{\sigma}_o^2[1, 5, 10, 15, 20, 100, 500, 1000, 2000, 4000] \end{aligned}$$

where $\hat{\sigma}_o^2$ is a robust estimate of the variance given by [16, p. 400]

$$\hat{\sigma}_o^2 = \left(\frac{\overline{\sigma}}{0.7}\right)^2 \tag{11}$$

where $\overline{\sigma}$ (the median absolute deviation) is the median of $|x(n) - \tilde{x}|$, and $\tilde{x}$ is the median of the observations $x(n)$. The final parameter

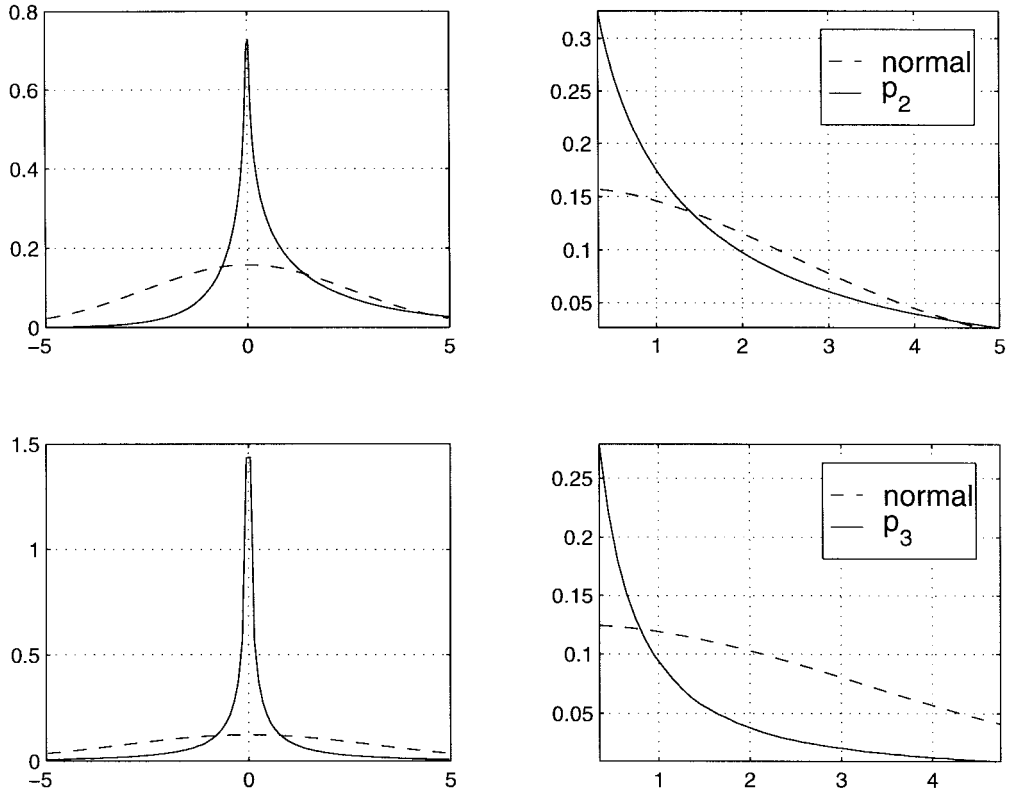

Fig. 3. Top: Theoretical first-order pdf of $p_2(n)$. Bottom: Estimated first-order pdf of $p_3(n)$. Normal pdf's, with variance matching that of $p_2(n)$ and $p_3(n)$, respectively, are shown for reference. The right-hand panels are enlargements of the left-hand ones.

TABLE IV
ESTIMATED PARAMETERS OF A TEN-TERM GAUSSIAN MIXTURE APPROXIMATION TO THE NON-GAUSSIAN PDF OF $p_3(n)$, FOR WHITE GAUSSIAN INPUT TO THE MZ-DWT

l	1	2	3	4	5	6	7	8	9	10
λ_l	0.0709	0.2429	0.2914	0.1562	0.0940	0.1107	0.0136	0.0102	0.0064	0.0036
σ_l^2	0.0000	0.0035	0.1031	0.8872	3.0992	14.0930	121.0198	128.3395	128.6320	128.6911

estimates of $\hat{f}_m(z)$ in (10) are shown in Table IV. The resulting Gauss-mixture pdf estimate compares well with the large-sample histogram. A slight overshoot near the origin and undershoot in the estimated pdf tail are typically observed, because a very large-sample size is needed to fully characterize the tails of a heavy-tailed distribution. However, we obtain useful estimates with as few as 100 samples. Note from Table IV that the last four terms are essentially identical, indicating that the mixture order L might be chosen smaller for this data set. Also note the $l = 1$ term results in a relatively sharp peak at the pdf origin.

Remark 1: We make some general comments on the first and second-order statistics (SOS) of $p(n)$, for input $x(n)$ a zero-mean stationary random process; further details may be found in [27] and [28]. It follows that $E[p_K(n)] = 0$ for K odd, if the odd-order cumulants of $x(n)$ are zero. In particular, $E[p_1(n)] = 0$ for $x(n)$ zero-mean, and $E[p_3(n)] = 0$ if $x(n)$ is zero-mean and has symmetric pdf such as Gaussian. However, in general, $E[p_K(n)] \neq 0$ for K even.

General expressions for the correlation $r_p(\tau) = E[p(n)p(n+\tau)]$ may be obtained in terms of the deterministic correlations of the impulse responses $h_j(n)$ and the moments of $x(n)$. For $x(n)$ white Gaussian, $p_K(n)$ is essentially uncorrelated, an effect that becomes more pronounced for increasing number of terms K in (4). The third- and fourth-order cumulants of $p_2(n)$ and $p_3(n)$ are also essentially delta functions at the origin, indicating that the multiscale products are essentially independent identically distributed (at least up to order four) for white-Gaussian input. We note that the case of $x(n)$ non-Gaussian can be handled in a similar manner by expressing the moments of $x(n)$ in terms of cumulants and exploiting the assumed i.i.d. nature of $x(n)$. This will also generally involve higher order analogs of the deterministic correlations of the $h_j(k)$.

The general case of $x(n)$ correlated can also be evaluated. For example, with $x(n)$ a Gaussian autoregressive (AR) (1) process, then $r_x(\tau) = r_x(0)(-a_1)^{|\tau|}$, with a_1 the AR coefficient. Now, the resulting products of $r_x(\tau)$ arising in $r_p(\tau)$ will result in $p(n)$ being a less correlated process than $x(n)$ because $|a_1| < 1$. Generally, this whitening effect is heightened as K increases. These correlation properties are beneficial in detection and estimation problems, and offset to some extent the heavy-tailed distribution.

Remark 2: In forming multiscale products, the individual outputs $y_j(n)$ could be raised to different powers before the product operation. Alternatively, one could search for minima in the reciprocal multiscale product. An interesting special case arises in products of the form

$$X_m = \prod_{j=1}^{m} (Y_j)^{-2j}. \tag{12}$$

For Y_j i.i.d. Gaussian, X_m is α-stable, with $\alpha = 2^{-m}$, and skewness parameter $\beta = 1$ [4]. With these values for α, X_m has finite moments only for orders less than 2^{-m}, which decreases rapidly

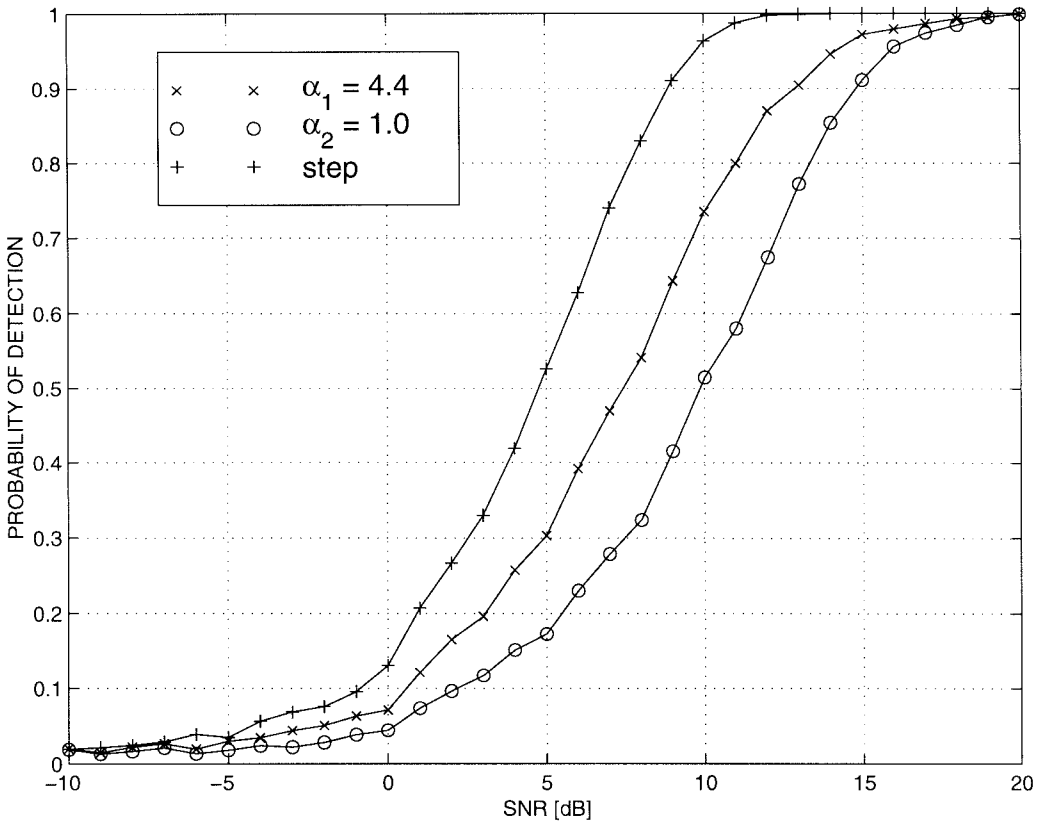

Fig. 4. Probability of step detection in white Gaussian noise using $|p_3(n)|$, with $P_{fa} = 0.01$. Results for three different edge shapes are shown (ideal step and sigmoidal with parameter α).

as m increases, resulting in very heavy-tailed distributions for X_m. The case of $m = 1$ is well known, corresponding to the Pearson V or Levy pdf. One could also use $\log X_m$. Applying this result to MZ-DWT outputs would require skipping scales in order to approximate the i.i.d. assumption; see Table II.

B. Detection in the Multiscale Product Domain

We consider a two-sided edge detection test based on $|p_3(n)|$ exceeding a threshold. Performance of this test is shown in Fig. 4 for three edge models. We used additive white-Gaussian noise with variance $\sigma_v^2 = 1$. The detection threshold was obtained by integrating the histogram estimate of the pdf described above, to achieve a probability of false alarm $P_{fa} = 0.01$ (the resulting threshold was 13.51). The predicted P_{fa} was verified via large-sample noise only Monte Carlo testing. The steps were of height $A = m_2 - m_1$, with SNR defined as

$$\mathrm{SNR} = 10\log_{10}\frac{A^2}{\sigma_v^2}. \tag{13}$$

One edge model was an ideal step with $m_1 = 0$. The other two edges were generated by sampling a continuous-sigmoidal step change signal given by (e.g., see Reza and Doroodchi [22])

$$\begin{aligned} x_s(n) &= \frac{m_2 + m_1 e^{-\alpha T(n-n_0-\tau)}}{1 + e^{-\alpha T(n-n_0-\tau)}} \\ &= \frac{A}{2}\left[1 + \tanh\left(\frac{\alpha T}{2}(n - n_o - \tau)\right)\right] + m_1, \\ &\qquad n = 0, \cdots, N-1. \end{aligned} \tag{14}$$

Here, T is the sampling interval time, parameter α determines the rise time, and the step occurs (in continuous time) at $n_0 T + \tau$. Without loss of generality we can assume $T = 1$, because T can always be incorporated into α, and we set the initial value $m_1 = 0$.

The phase parameter τ was uniformly distributed in $[0, T]$, modeling the effects of time quantization via sampling. The record length was $N = 256$ and $n_0 = N/2$. For the two cases of (14) we used $\alpha = 4.4$, and $\alpha = 1$, corresponding to a fast and slower step rise-time, respectively. In all cases m_2 was adjusted to achieve the desired SNR. Five-thousand Monte Carlo trials were run for each SNR value, and $|p_3(n)|$ was tested at the position where the edge should produce a maximum. As we would expect, slower rise-time results in poorer detection performance. The ideal step results compare favorably with linear operators; e.g., see Abdou and Pratt [1].

V. Estimation of Step Change Location

Next we consider estimation of step change location. This assumes that a change has occurred in the observation interval, hence there is no detection problem. We have developed a general closed-form CRB for step-change estimation, valid for additive-correlated Gaussian and i.i.d. non-Gaussian noise, as well as specific expressions for the sigmoidal step-change model given by (14) [31], [28], [27]. These results have also been generalized to the additive-multiplicative noise case [31], [30], see also [8], [11], and [29]. Let the noisy discrete-time signal be given by

$$\begin{aligned} y(n) = x(n) + v(n) = s(\alpha(n - n_o - \tau)) + v(n), \\ n = 0, \cdots, N-1 \end{aligned} \tag{15}$$

where $v(n)$ is zero-mean i.i.d. noise with variance σ_v^2. The step-change occurs at time $\rho = n_o + \tau$, where τ is the sampling phase, and α is a scaling parameter that determines the rise-time of the step (see the previous section). For τ random and uniformly distributed

over the sampling interval, a general *closed-form* expression for the Fisher information for ρ is given by

$$J = \alpha \frac{\gamma_{v0}}{\sigma_v^2}[G(\alpha(N - n_0 - 1)) - G(\alpha(-n_0 - 1))] \tag{16}$$

where

$$G(u) = \int [s'(u)]^2 \, du \tag{17}$$

and

$$\gamma_{v0} := \sigma_v^2 \int \left[\frac{dp_v(u)}{du}\right]^2 \frac{1}{p_v(u)} \, du = \sigma_v^2 I_p. \tag{18}$$

In (18) I_p is the Fisher Information for the location parameter of the pdf $p_v(v)$. For any reasonable step-approximator, $s(u)$, its derivative will vanish outside a more or less small interval centered around n_0, and we expect that $G(u)$ will attain its asymptotic values rather quickly. The asymptotic value is obtained by assuming that N is large and that n_0 is not too close to either end of the observation window, yielding

$$J_\infty = \alpha \frac{\gamma_{v0}}{\sigma_v^2}[G(+\infty) - G(-\infty)]. \tag{19}$$

For the sigmoidal step-model of (14) we find the closed-form Fisher information to be

$$\begin{aligned} J = \frac{(\alpha T) S \gamma_{v0}}{8} \Big[& \tanh(\alpha T(N - n_0 - 1)/2) \\ & - \tanh(\alpha T(-n_0 - 1)/2) \\ & - \frac{1}{3}\tanh^3(\alpha T(N - n_0 - 1)/2) \\ & + \frac{1}{3}\tanh^3(\alpha T(-n_0 - 1)/2) \Big] \end{aligned} \tag{20}$$

where $S \triangleq (m_2 - m_1)^2/\sigma_v^2 = A^2/\sigma_v^2$ is the SNR. An asymptotic form of (20) is found to be

$$J_\infty = \frac{\alpha T S \gamma_{v0}}{8} \frac{4}{3} = \frac{\alpha T S \gamma_{v0}}{6}. \tag{21}$$

Evaluation of (20) and (21) shows weak dependence on d, the fractional index of the location parameter ($d \approx 0.5$ corresponds to the step-change occurring approximately at the center of the observation window). The Fisher information reaches its asymptotic maximum for a range of d of approximately $0.2 \leq d \leq 0.8$, for $\alpha T N > 15$. Even for $\alpha T N = 2$ the Fisher information is about half its asymptotic value. This indicates that only local information around the step-change is necessary to achieve good estimation of location.

It is also interesting to consider the case when τ is a nonrandom fixed constant τ_0. The Fisher information for this case is given by

$$J = \frac{(\alpha T)^2 \gamma_{v0}}{16} \frac{A^2}{\sigma_v^2} \sum_{k=0}^{N-1} \cosh^{-4}(\alpha T(k - n_0 - \tau)/2) \tag{22}$$

where $A = m_2 - m_1$ denotes the step-height. We plot this case in Fig. 5, showing Fisher information versus parameter αT, with curves parameterized by $0 \leq \tau \leq 0.5$. Here, $N = 10$, and $\gamma_{v0} A^2/16\sigma_v^2$ was set to unity. Due to symmetry the case of $\tau = .6$ is equivalent to that for $\tau = 0.4$, and so on. The case of $\tau = 0$ (or $\tau = 1$) corresponds to the step-change being centered exactly at the sample time, so that the Fisher information increases without bound as αT increases. The curves for $\tau \neq 0$ in Fig. 5 show the sensitivity to τ in estimating step-location. Depending on the step rise time and the sampling time, it matters where the samples fall on the continuous-representation of the step. We note that as $\alpha T \to \infty$, the step-rise time becomes instantaneous, and all of the curves decrease to zero in the limit (except the $\tau = 0$ and $\tau = 1$ cases). In this limit the samples will bracket an instantaneous step-change, and no subsample time accuracy can be determined unless τ is known.

As an example we show in Fig. 6 the theoretical CRB given by the inverse of (20), as well as experimental mean-square error (MSE), for two edge location estimation methods. The step was modeled via (14) with τ uniformly distributed. The two estimation methods are first, based on $p_3(n)$, and second, based on a simple gradient estimator with FIR given by [−1, 0, 1]. The sigmoidal function was generated with $m_1 = 0$, $T = 1$, and $\alpha = 4.4$, corresponding to a rise-time of about $4T$ (see [22]). The step-height m_2 was set to achieve the desired SNR. We used $N = 256$ and averaged over 5000 Monte Carlo trials for each SNR value. For each trial τ was selected uniformly in [−0.5, 0.5], and the step was centered at $N/2$. For the simple gradient, the step location was taken as the maximum in a 32 sample-observation window centered on $N/2$. In the noise-free case with $\alpha = 4.4$, the sampled sigmoidal step-rise occupies two samples, leading to an ideal $p_3(n)$ response that is a two-sample wide plateau. The step-location was taken as the center of this ideally two-pixel wide plateau, hence the peak location estimate is quantized to $T/2 = 1/2$. This results in both methods being unbiased at high SNR with $\tau = 0$. Note that the CRB in Fig. 6 is for *continuous* parameter estimation, whereas the estimators are discrete. Thus, with τ randomized as in Fig. 6, both methods become quantization error limited at high SNR; further improvement toward the CRB requires sub-pixel interpolation.

Changing α changes the effective rise-time of the sigmoidal signal, and smaller α creates a slower rising step. This results in a shift upward of the CRB as the step-location is inherently more difficult to estimate. Repeating the experiment in Fig. 6 with lower α results in a shift to the right as higher SNR is needed to obtain the same performance, while the general behavior of the curves is the same.

VI. Discussion

Multiscale products may be used to create a delta-train waveform that aligns with edge transitions as in Fig. 1, employing a memoryless-product nonlinearity, and the method is of low complexity. The products tend to reduce input noise correlation while producing a very local response to step transitions, desirable properties in this application. An odd number of products may be generally prefered, as the result is unbiased and preserves edge-polarity information. We have focused on the first few dyadic scales, higher scales may be employed depending on application. As the region of support grows with scale (see Table I), closely neighboring edges will further interact and be suppressed (see Fig. 2). However, under-smoothing must be avoided or spurious peaks will result. More closely spaced nondyadic scales may also be employed, however this increases the degree of correlation of the noise across scales, and dyadic scales seems to be a good general choice.

With regard to the general step-change problem, we have developed closed-form CRB expressions for the discrete-time case that may be readily evaluated for specific underlying continuous-time edge models. We have shown results for a sigmoidal model, and illustrated the effects of sampling phase. The CRB is conveniently formulated in terms of a rise-time parameter, the sampling rate, the SNR, and the sampling phase. CRB results underline the fundamental tradeoff between detection (many samples desired) and estimation (few samples required if the edge is known to occur in a neighborhood). In practice discrete-time methods become quantization limited at high SNR, i.e., they are unable to continue to resolve the time of the edge occurrence because of the finite sampling rate. Interpolation or other subsample schemes can produce more accurate results, but care must be taken at low SNR as interpolation may make things worse.

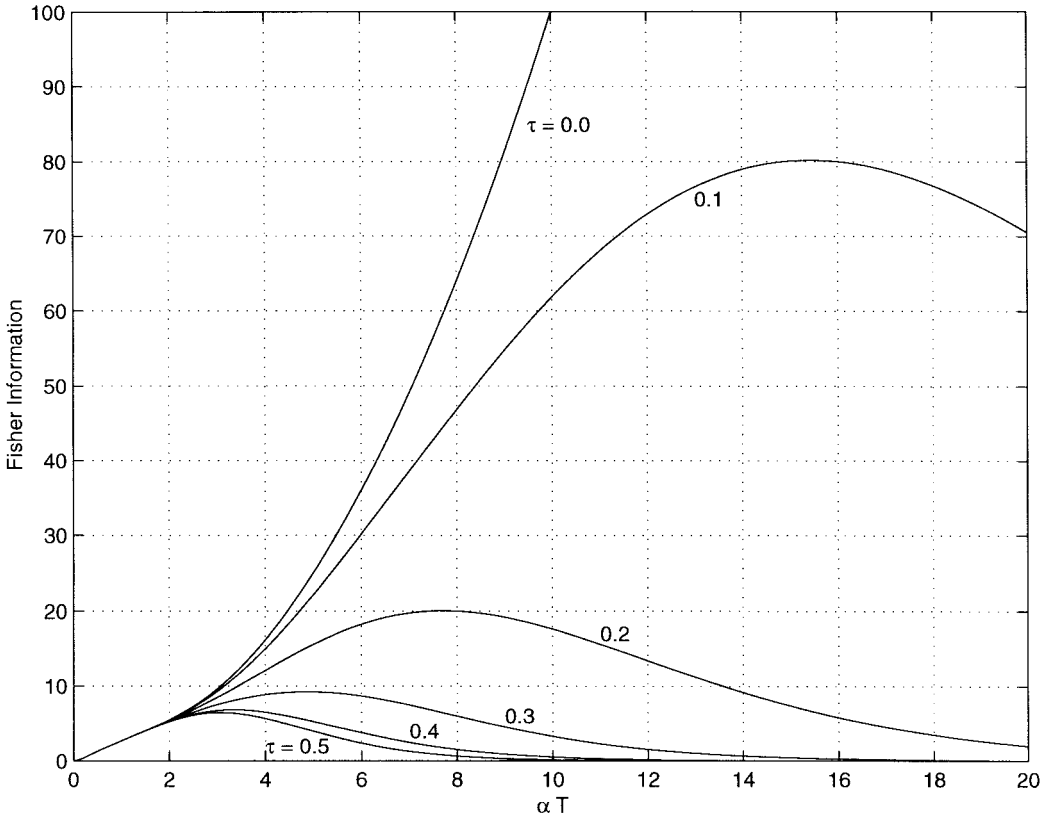

Fig. 5. Fisher information for step-change location estimation with sample-time offset τ nonrandom. Curves parameterized by τ; $\tau = 0$ implies sample time and step-change aligned; $\tau = 0.5$ implies step-change center exactly between samples.

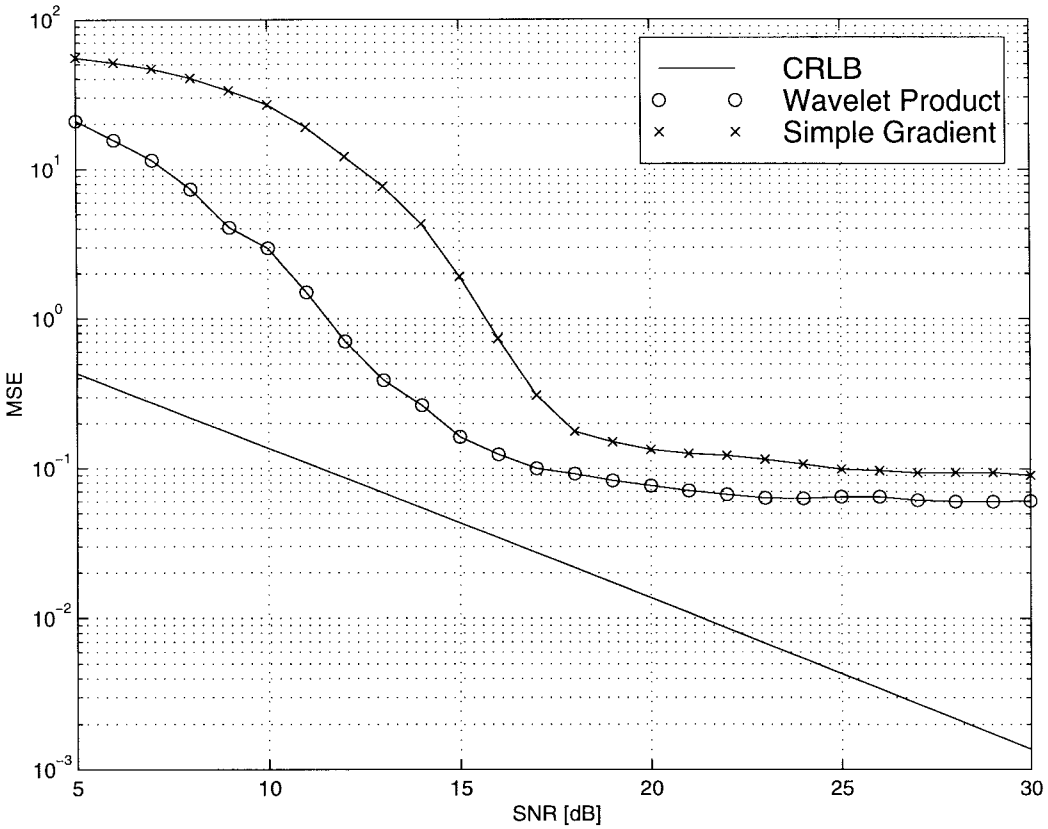

Fig. 6. MSE's and theoretical CRB for step change location estimation.

Many nonlinear methods have been proposed for constrained locally adaptive smoothing, such as local monotonicity [21], regularization [9], and wavelet denoising. These filters may be followed by multiscale-gradient estimation for edge detection, and represent an interesting higher complexity alternative to the linear smoothing filters studied here.

ACKNOWLEDGMENT

The authors are grateful to R. Kozick for his assistance with the EM algorithm, and to F. Gini for material related to the CRB's.

REFERENCES

[1] I. E. Abdou and W. K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," *Proc. IEEE*, vol. 67, no. 5, pp. 753–763, 1979.
[2] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes, Theory and Application.* Englewood Cliffs; NJ: Prentice-Hall, 1993.
[3] R. S. Blum, R. J. Kozick, and B. M. Sadler, "An adaptive spatial diversity receiver for non-Gaussian interference and noise," *IEEE Trans. Signal Processing*, to be published..
[4] A. W. Brown and T. J. Tukey, "Some distributions of sample means," *Annals of Math. Stat.*, vol. 17, no. 1, pp. 1–12, 1946.
[5] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679–698, Nov. 1986.
[6] Z. Cvetkovic and M. Vetterli, "Discrete-time wavelet extreme representation: Design and consistent reconstruction," *IEEE Trans. Signal Processing*, vol. 43, pp. 681–693, Mar. 1995.
[7] A. Denjean and F. Castanie, "Mean value jump detection: A survey of conventional and wavelet based methods," in *Wavelets: Theory, Algorithms, and Applications*, C. Chui, L. Montefusco, and L. Puccio, Eds. Newkoski Academic, 1994.
[8] M. Ghogho and A. Swami, "Fast computation of the exact FIM for deterministic signals in colored noise," *IEEE Trans. Signal Processing*, vol. 47, pp. 52–61, Jan. 1999.
[9] M. Gokmen and C. C. Li, "Edge detection and surface reconstruction using refined regularization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 5, pp. 492–499, 1993.
[10] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[11] R. Kakarala and A. O. Hero, "On achievable accuracy in edge localization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 777–781, July 1992.
[12] J. Koplowitz and V. Greco, "On the edge location error for local maximum and zero-crossing edge detectors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 12, pp. 1207–1212, 1994.
[13] R. Kozick, R. Blum, and B. M. Sadler, "Signal processing in non-Gaussian noise using mixture distributions and the EM algorithm," *in 31st Asilomar Conf. on Sig., Syst., and Comp.*, Pacific Grove, CA, Nov. 1997, pp. 438–442.
[14] C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 21–28, 1995.
[15] L. A. Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 5, pp. 729–734, 1982.
[16] L. Ljung, *System Identification: Theory for the User* Englewood Cliffs, NJ: Prentice-Hall, 1987.
[17] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 617–643, 1992.
[18] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 710–732, July 1992.
[19] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[20] K. S. Miller, *Multidimensional Gaussian Distributions*. New York: Wiley, 1964.
[21] A. Restrepo and A. C. Bovik, "Locally monotonic regression," *IEEE Trans. Signal. Processing.*, vol. 41, pp. 2796–2810, 1993.
[22] A. M. Reza and M. Doroodchi, "Cramer-Rao lower bound on locations of sudden changes in a steplike signal," *IEEE Trans. Signal Processing*, vol. 44, no. 10, pp. 2551–2556, 1996.
[23] A. Rosenfeld, "A nonlinear edge detection technique," *Proc. IEEE* pp. 814–816, May 1970.
[24] A. Rosenfeld, Y. H. Lee, and R. B. Thomas, "Edge and curve detection for texture discrimination," in *Picture Processing and Psychopictorics*. New York: Academic, 1970, pp. 381–393.
[25] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. 20, pp. 562–569, May 1971.
[26] B. M. Sadler, T. Pham, and L. C. Sadler, "Optimal and wavelet-based shockwave detection and estimation," *J. Acoust. Soc. Amer.*, vol. 104, no. 2, pt. 1, pp. 955–963, Aug. 1998.
[27] B. M. Sadler and A. Swami, "On multiscale wavelet analysis for step estimation," *in Int. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP'98), Seattle, WA, vol. 3, pp. 1517–1520, Apr. 1998.
[28] ——, "Analysis of wavelet transform multiscale products for step detection and estimation," Army Research Laboratory Tech. Rep., ARL-TR-1664, 1998.
[29] A. Swami, "Cramer-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Processing*, vol. 53, pp. 231–244, 1996.
[30] A. Swami and B. M. Sadler, "Step-change localization in additive and multiplicative noise via multiscale products," *in 32nd Asilomar Conf. on Sig., Syst., and Comp.*, Pacific Grove, CA, Nov. 1998.
[31] ——, "Cramer-Rao bounds for step-change localization in additive and multiplicative noise," *in 9th IEEE Statistical Signal and Array Processing Workshop*, Portland, OR, Sept. 1998, pp. 403–406.
[32] H. D. Tagare and R. J. P. deFigueiredo, "On the localization performance measure and optimal edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 12, pp. 1186–1190, 1990.
[33] Y. Xu, J. Weaver, D. Healy, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Trans. Image. Processing*, vol. 3, pp. 747–758, Nov. 1994.

Information Measures in Scale-Spaces

Jon Sporring and Joachim Weickert

***Abstract*— This correspondence investigates Rényi's generalized entropies under linear and nonlinear scale-space evolutions of images. Scale-spaces are useful computer vision concepts for both scale analysis and image restoration. We regard images as densities and prove monotony and smoothness properties for the generalized entropies. The scale-space extended generalized entropies are applied to global scale selection and size estimations. Finally, we introduce an entropy-based fingerprint description for textures.**

***Index Terms*—Scale-space, Shannon–Wiener entropy, Rényi's generalized entropies, scale selection, size estimation, texture.**

I. INTRODUCTION

In recent years multiscale techniques have gained a lot of attention in the image processing community. Typical examples are pyramid and wavelet decompositions. They represent images at a small

Manuscript received February 10, 1998; revised October 15, 1998. This work was funded in part by the Real World Computing Partnership, the Danish National Research Council, and the EU-TMR Project VIRGO.

J. Sporring was with the Image Group, Department of Computer Science, University of Copenhagen, DK-2100 Copenhagen, Denmark. He is now with the Foundation for Research and Technology–Hellas (FORTH), GR 711 10 Heraklion, Crete, Greece.

J. Weickert is with the Image Group, Department of Computer Science, University of Copenhagen, DK-2100 Copenhagen, Denmark.

Publisher Item Identifier S 0018-9448(99)02269-5.